# Air passengers 1949-1960

**22 marks**

**Airline passengers**.
Here we will examine the monthly total of the number of air passengers in the US from 1949 to 1960.

The data are available as the data set `AirPassengers` from the `datasets` package in the standard `R` distribution.

```
AirPassengers
```

```
##      Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
## 1949 112 118 132 129 121 135 148 148 136 119 104 118
## 1950 115 126 141 135 125 149 170 170 158 133 114 140
## 1951 145 150 178 163 172 178 199 199 184 162 146 166
## 1952 171 180 193 181 183 218 230 242 209 191 172 194
## 1953 196 196 236 235 229 243 264 272 237 211 180 201
## 1954 204 188 235 227 234 264 302 293 259 229 203 229
## 1955 242 233 267 269 270 315 364 347 312 274 237 278
## 1956 284 277 317 313 318 374 413 405 355 306 271 306
## 1957 315 301 356 348 355 422 465 467 404 347 305 336
## 1958 340 318 362 348 363 435 491 505 404 359 310 337
## 1959 360 342 406 396 420 472 548 559 463 407 362 405
## 1960 417 391 419 461 472 535 622 606 508 461 390 432
```

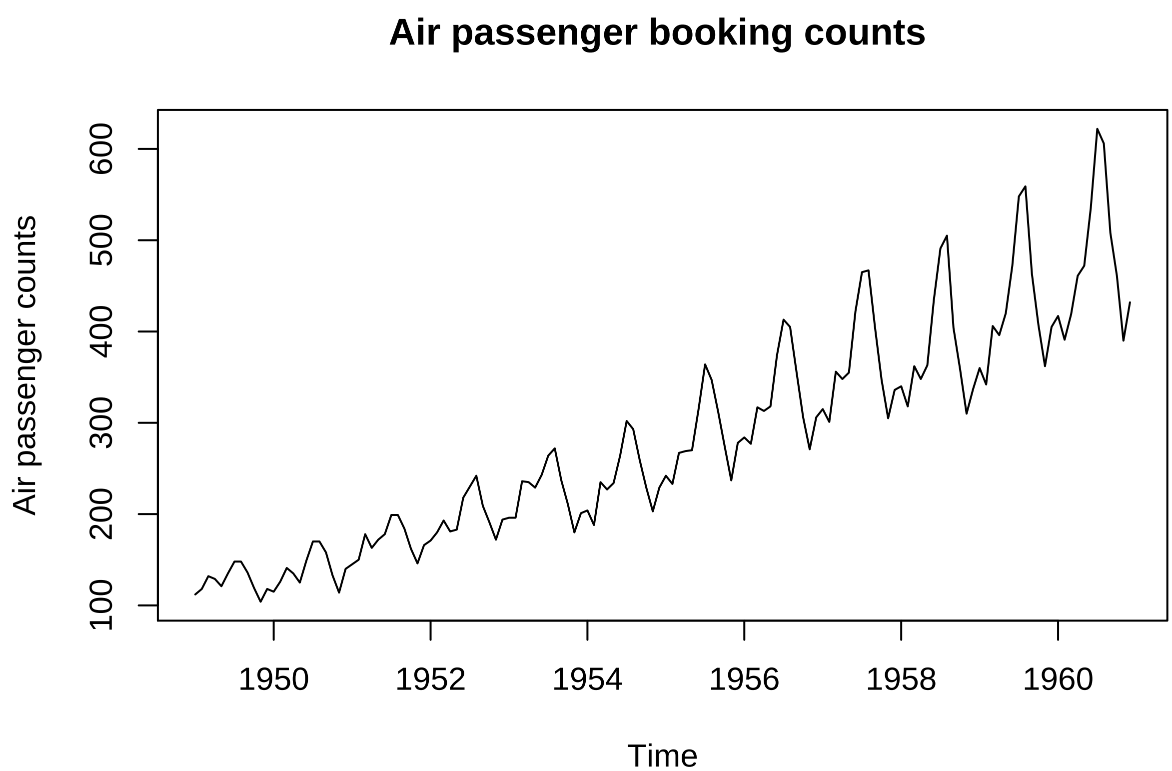
a. **(7 marks)** Use the function `decompose()` on this data and plot the results.

   i. *(2 marks)* Show your plot.
   ii. *(1 mark)* Describe the trend.
   iii. *(2 marks)* Describe the seasonal pattern.
   iv. *(2 marks)* What do you conclude from the residuals?

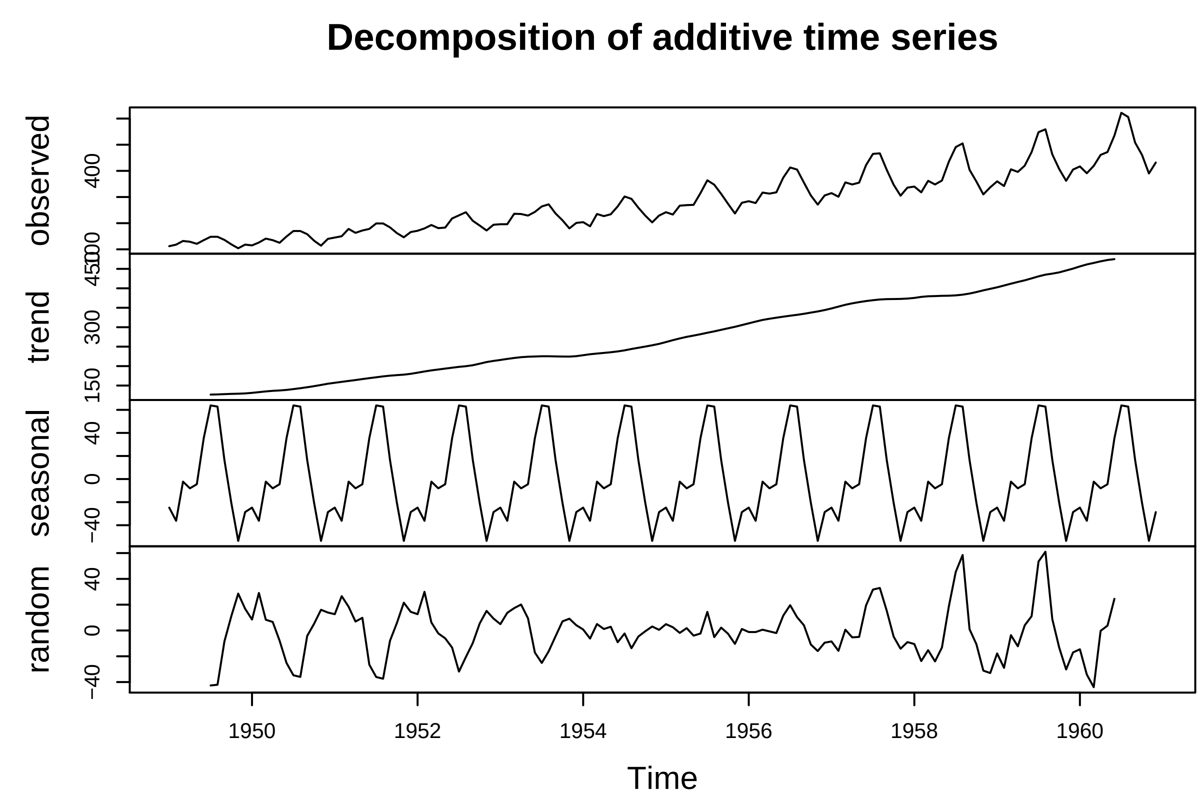
```
AP = AirPassengers
frequency(AP)
```

```
## [1] 12
```

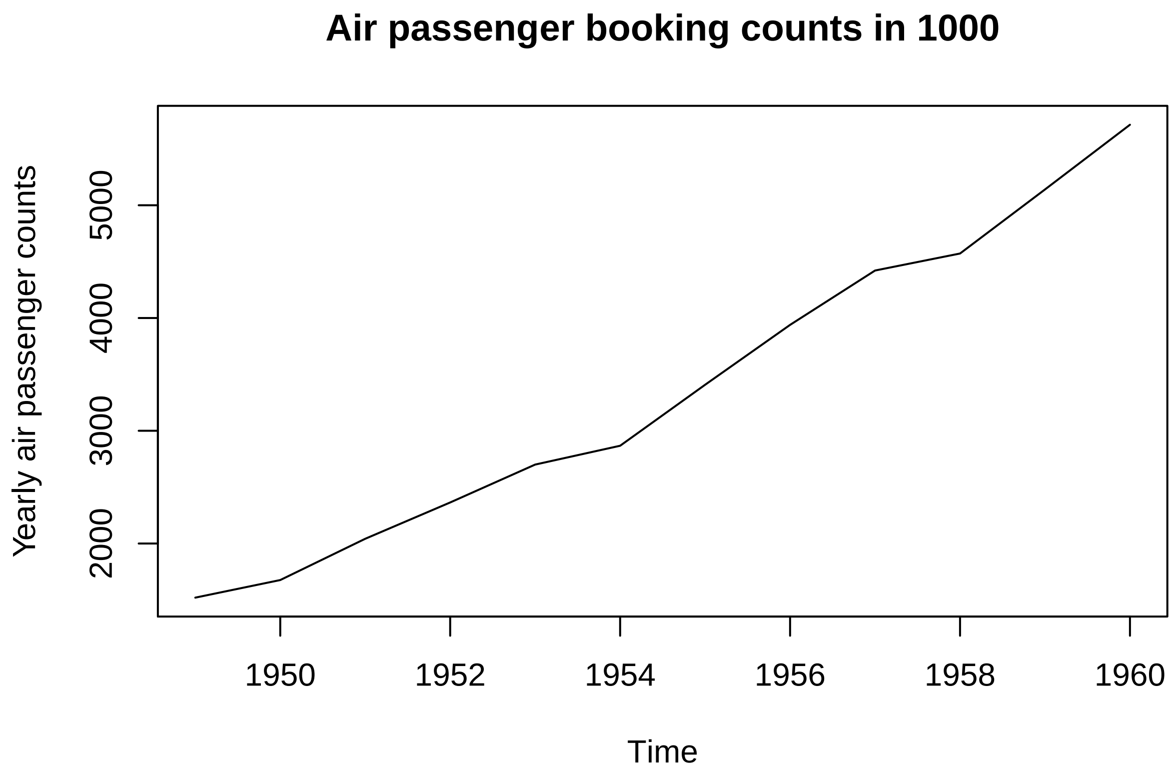
```
plot(AP, ylab="Air passenger counts", main="Air passenger booking counts ")
```



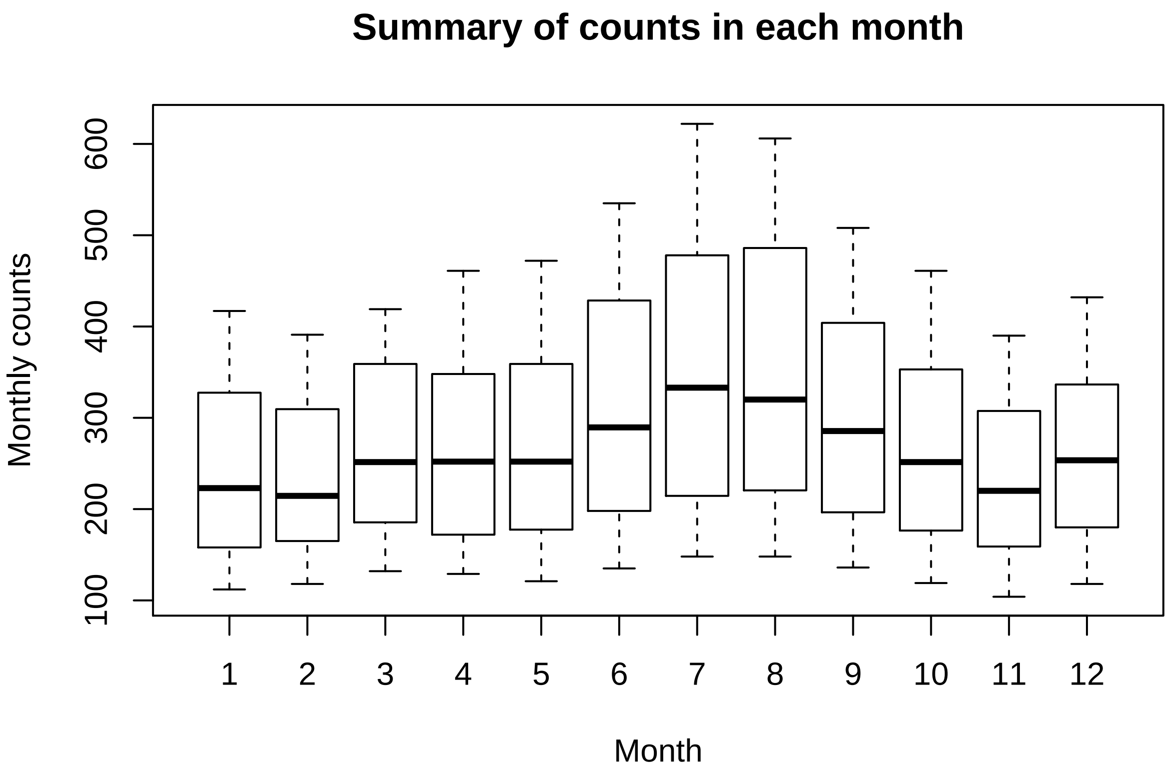

```r
plot(decompose(AP))
```



```r
trend.AP <- aggregate(AP)
plot(trend.AP, ylab="Yearly air passenger counts", main="Air passenger booking counts in 1000")
```




```r
boxplot(AP~cycle(AP), xlab="Month", ylab="Monthly counts", main="Summary of counts in each month")
```




(ii) As the data and the frequency is 12, we can see the data is collected monthly. There is a clear upward trend shown in the decomopsotion graph. As the time increases, the mean of the data increases.

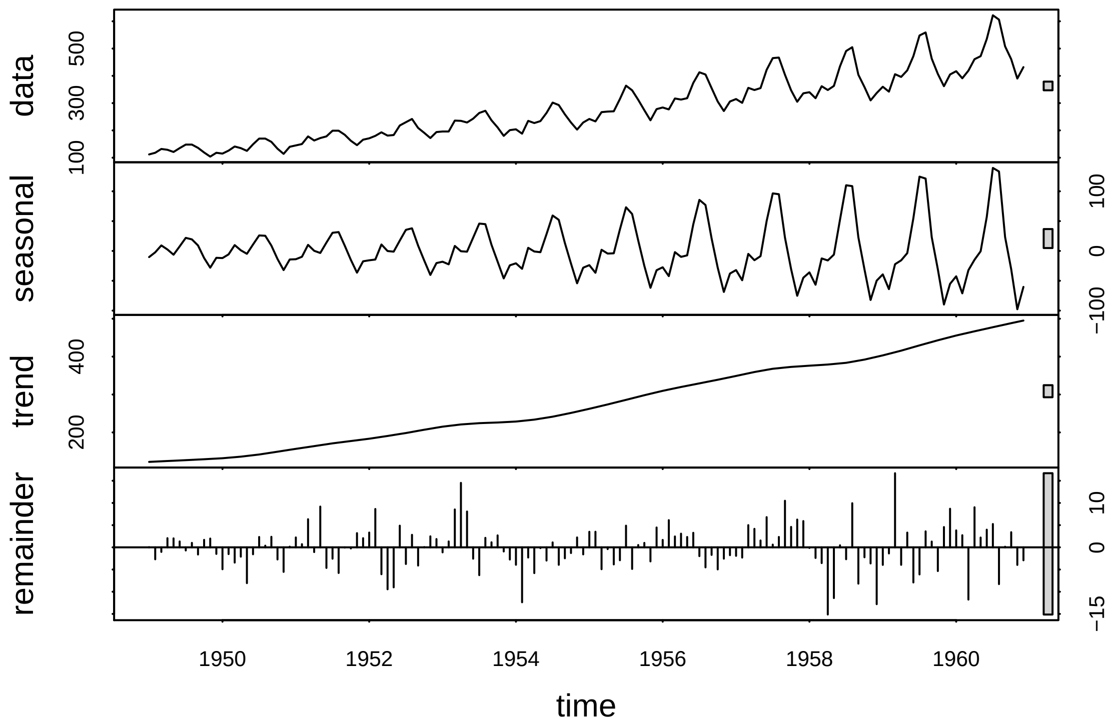
(iii) Also there is a constant and obvious seasonality shown, more people travel in the middle of the year, which is June to September. That implies that each year have the same peak and low season effect.

(iv) From the residual, there are 2 change points, one is around 1954 and the other is around 1957. After the change point, the pattern of the residuals changes. Overall, it has mean of zero for the residuals. The variability is much more constant 1954 to 1957, others are more flucate.

b. **(9 marks)** Use the function `stl()` on this data with seasonality loess span `s.window = 7` and the local polynomial for the seasonal loess being a line.

i. *(2 marks)* Show your plot.
ii. *(2 marks)* Describe the trend. Is it significant? Why or why not?
iii. *(3 marks)* Describe the seasonal pattern. Is it significant? Why or why not?
iv. *(2 marks)* What do you conclude from the residuals?

```
plot(stl(AP, s.window = 7, s.degree = 1))
```




(ii) There is an obvious upward trend shown in the graph. As the grey small box shown in the right is quite small for the trend graph, it implies that the trend is quite big, therefore the trend is significant.

(iii) For the seasonality, even if the seasoanlity is changing with time, the grey box on the right still small, thus the seasonality is still significant; however, with comparsion of the grey box, it is shown that the seasonality is less significant than the trend. As the seasonality pattern increases with time, it shows that the effect on peak and low season increases as time goes.

(iv) There is not clear change point shown by the residual graph. The residual vallue is quite small (also use the grey box as a reference size), and it has an approximate mean of 0, therefore it is not significant.

c. **(6 marks)** Compare the plots from parts (a) and (b).

i. *(2 marks)* What are the major differences in the plots?

As the function we used are different, the seasonality shown is the major difference. In the decomposition graph, the seasonality is more as a non-changable variable for all years with mean of zero. For the stl plot, the sesoanlity increases as time increases still with a zero mean.

ii. *(2 marks)* What characteristics of the two methods caused these differences?

In the decomposition graph, we use the average over years for each month and then use the average for that month ever year, in stl we smooth the values as for the estimate for that month for that year. With that, the seasonality value in decomposition should always be the same as it uses the average. For stl, aw they try to smooth the value rather than taking average, it is more related to the observed value. As the amplitude for the observed value increases, the stl seasonality with the smoothing value will increase its amplitude, therefore the the seasonality are shown differently.

iii. *(2 marks)* Which fitted model do you prefer and **why**?

I prefer the stl. As the observed data shown, the data is not just increase overtime but the amplitude for each year also increases with time. With respect to it, a constant seasonality each year can not explain this change, thus the stl with the increasing size of amplitude of seasonality is more fittable for the data.